Accepted Manuscript

Descriptive Analysis of the Verbal Behavior of a Therapist: A Known-Group Validity Analysis of the Putative Behavioral Functions Involved in Clinical Interaction

Javier Virues-Ortega, Montserrat Montaño-Fidalgo, María Xesús Froján-Parga, Ana Calero-Elvira

PII: S0005-7894(11)00047-5
DOI: doi: 10.1016/j.beth.2010.12.004
Reference: BETH 260

To appear in: *Behavior Therapy*

Received date: 17 May 2010
Revised date: 5 December 2010
Accepted date: 30 December 2010

Please cite this article as: Virues-Ortega, J., Montaño-Fidalgo, M., Froján-Parga, M.X. & Calero-Elvira, A., Descriptive Analysis of the Verbal Behavior of a Therapist: A Known-Group Validity Analysis of the Putative Behavioral Functions Involved in Clinical Interaction, *Behavior Therapy* (2011), doi: 10.1016/j.beth.2010.12.004

Descriptive Analysis of the Verbal Behavior of a Therapist: A Known-Group Validity Analysis of the Putative Behavioral Functions Involved in Clinical Interaction

Javier Virues-Ortega

University of Manitoba and St. Amant Research Centre

Montserrat Montaño-Fidalgo

María Xesús Froján-Parga

Ana Calero-Elvira

Universidad Autónoma de Madrid

The authors would like to thank Dr. Joseph Pear (University of Manitoba) for his comments on an earlier version of this manuscript.

Address correspondence to Javier Virues-Ortega, Ph.D., Psychology Department, University of Manitoba, 177 Dysart Road, MB R3T Winnipeg, Manitoba, Canada; e-mail: javier_virues@umanitoba.ca.

Abstract

This study analyzes the interobserver agreement and hypothesis-based known-group validity of the Therapist's Verbal Behavior Category System (SISC-INTER). The SISC-INTER is a behavioral observation protocol comprised of a set of verbal categories representing putative behavioral functions of the in-session verbal behavior of a therapist (e.g., discriminative, reinforcing, punishing, and motivational operations). The complete therapeutic process of a clinical case of an individual with marital problems was recorded (10 sessions, 8 hours), and data were arranged in a temporal sequence using 10-min periods. Hypotheses based on the expected performance of the putative behavioral functions portrayed by the SISC-INTER codes across prevalent clinical activities (i.e., assessing, explaining, Socratic method, providing clinical guidance) were tested using autoregressive integrated moving average (ARIMA) models. Known-group validity analyses provided support to all hypotheses. The SISC-INTER may be a useful tool to describe therapist–client interaction in operant terms. The utility of reliable and valid protocols for the descriptive analysis of clinical practice in terms of verbal behavior is discussed.

Over the last few decades, the clinical psychology literature has emphasized outcome research over research on the molecular processes occurring during clinical interaction (Lambert & Hill, 1994). Researchers have focused both on the overall effectiveness of specific interventions and the comparative effectiveness across paradigms rather than the effects of within-paradigm molecular processes. Moreover, the ubiquitous implementation of group-based methodology in clinical psychology, and more specifically, clinical trials, may have been detrimental to the study of molecular clinical processes, which may be better suited to a single-subject approach.

The development of valid and reliable strategies for observing, coding, and analyzing language during clinical interaction is a methodological prerequisite to an in-depth analysis of in-session clinical processes. Various observational protocols from various clinical psychology paradigms have been developed with that aim (e.g., Simon & Agazarian, 2000), but there is a paucity of behavioral literature in this area. From a behavior-analytic standpoint, W. Day developed a method to analyze transcripts of real clinical sessions in order to identify specific verbal operants and corresponding antecedents and consequences (unpublished work cited in Moore, 1991). Such an approach could be construed as qualitative in that it relied heavily on the observer's theoretical knowledge and his or her interpretation of categorizations of verbal behavior (Moore, 1991). This form of analysis of in-session verbal behavior has had a low impact on clinical research (Hayes et al., 2001; Moore, 1991) and was limited in several crucial ways: (a) it was highly time consuming, (b) observation codes varied across sessions and clients impeding comparability, and (c) there was no way to systematically assess interobserver agreement and validity (Dougher, 1989).

Conducting a molecular analysis of in-session clinical interaction as the basis for explaining clinical phenomena from a behavioral standpoint still has significant heuristic value. Borckardt et al. (2008) have shown that time-series recording of in-session clinical events is a methodologically sound method for studying causal therapist–client interactive processes in psychotherapy and psychopathology (see also Jones, Ghannam, Nigg, & Dyer, 1993; Smith, Wolf, Handler, & Nash, 2009). In addition, behavior analysis seems to be a particularly well-suited approach to attempt the analysis of the molecular processes of psychotherapy because (a) clinical interaction could be analyzed in terms of putative contingencies (antecedent–behavior–consequence sequences), (b) there is abundant single-subject experimental literature showing how different behavior processes operate, and (c) there is a conceptual basis for behavior change related to verbal behavior analysis (e.g., Catania, 2006).

Busch et al. (2009), using the Functional Analytic Psychotherapy Rating Scale, showed that shaping responses was more likely to occur during functional–analytic psychotherapy as opposed to cognitive–behavioral therapy. Observation codes in that study were largely based on the conceptual underpinnings of functional–analytic psychotherapy (e.g., “therapist evokes clinically relevant behavior,”) and relied on the observers’ judgment to determine whether a specific behavior process (e.g., shaping) had occurred. Focusing observation on more basic behavioral functions and incorporating methodological strategies to test specific hypotheses on the expected performances of those functions could enhance Busch et al.’s analysis. For example, an observation system suggesting that a therapist’s utterance (e.g., instances of therapist’s approval) was associated with increases in temporally related client behaviors, would indicate that a reinforcement effect is likely to have occurred. An approach based on the indirect identification of putative behavior functions could have a higher potential for (a) identifying behavioral processes more objectively (as opposed to relying on an observer’s judgment to

determine their occurrence), (b) developing hypotheses on the active components in any model of psychotherapy, and (c) being used as the basis for monitoring the therapeutic process for clinical decision making and quality assurance purposes.

Observation codes were developed in accordance with the putative behavioral functions of verbal utterances in terms of operant theory (e.g., basic behavioral principles such as reinforcement and punishment) and used for recording and analyzing a therapist's (clinical) verbal behavior; using such an assessment protocol allows for the assessment of interobserver agreement. A methodological approach to assessing the construct validity of such a descriptive analysis is introduced and discussed in the context of a therapist's clinical interaction with a client, from the first assessment session through to the last follow-up session. Therefore, what follows is a methodological study presenting a construct validity strategy and not an experimental study on the specific behavior processes that will be hypothesized in order to inform validity.

Method

PARTICIPANTS

The participant was a 32-year-old woman who had been seeking help for her marital problems. Clinical sessions took place in a private clinic and the participant had voluntarily approached the clinic for therapeutic input. The therapist was a 44-year-old female and licensed clinical psychologist with a long history of providing behavior therapy. Both the therapist and the participant were aware that the sessions were to be recorded for research purposes but they were not familiar with the specific goals and hypotheses of the study. The research protocol was approved by the Human Subjects Committee of the Universidad Autónoma de Madrid (Spain) and the participant, therapist, and clinic director all provided signed consent for their involvement in the study. Confidentiality and privacy was maintained throughout.

Intervention Description

The therapist implemented cognitive–behavior therapy (see Epstein & Baucom, 2002, for a general approach to assessment and treatment). The first three sessions concentrated on assessment. In the fourth session the therapist presented her conceptualization of the case to the participant along with a description of the intervention plan. During the next sessions the treatment plan was implemented. Intervention was discontinued when treatment goals were fully met (Session 10). The participant's partner participated in only one of the assessment sessions. The first six sessions occurred weekly, Sessions 7–9 occurred biweekly, and Session 10 occurred 1 month after the previous session. Qualitative descriptions of the major and minor areas covered in each session are summarized in Table 1.

Observation Protocol

The Therapist's Verbal Behavior Category System (SISC-INTER) was used to record and categorize the verbal behavior of the therapist (Froján et al., 2008). The SISC-INTER categories were developed by a panel of senior behavior therapists based on the likely behavioral functions of the therapist's utterances during clinical interaction. Informal observations were used as the basis for deriving observation codes that were later trialed until a systematic and exhaustive set of codes that were thought to capture the putative verbal behavior functions was obtained. The following SISC-INTER codes were used in this study to categorize specific functions of verbal behavior:

- Discriminative (Di): therapist verbalization followed by a specific verbal or nonverbal response by the client. "Specific" meaning that client's response is clearly dependent, in

terms of content and temporal contiguity, on the therapist's preceding verbalization (e.g., Therapist: "When did the problem start?" [Di]; Client: "Two years ago more or less").

- Reinforcing (Rf): therapist verbalization denoting agreement, approval, or acceptance of a specific client's behavior (e.g., Client: "I went out last Friday as you suggested and I had a good time"; Therapist: "That is great! I am glad to hear it" [Rf]).
- Punishing (Pn): therapist verbalization denoting disapproval, refusal, or lack of acceptance of client's behavior (including interrupting the client without signs of approval, agreement, or acceptance; e.g., Therapist: "You have been offered a better position. I take it they regard you very highly, right?"; Client: "Nothing out of the ordinary I guess"; Therapist: "Come on, you have been offered the best position. I doubt that everybody would get that treatment" [Pn]).
- Informative (If): therapist verbalization conveying technical or clinical information in a plain-language format (e.g., Therapist: "That is exactly what happens when someone systematically avoids a feared situation. Emotions get more and more intense and unpleasant toward that situation and coping becomes unbearable. That is the point where you are at").
- Motivational (Mt): therapist verbalization anticipating the consequences, whether positive or negative, of a client's behavior (e.g., Therapist: "If you would stop answering his calls, he would stop calling eventually").
- Instructional (It): therapist verbalization describing specific directions/instructions for the client to be followed outside the session (e.g., Therapist: "As soon as you find yourself experiencing those feelings you just described, you should follow the simple steps for the relaxation exercise we have talked about before: Step 1, Step 2, etc.").

- Other (O): therapist's verbalization not included in the categories above (e.g., Therapist: "So you were saying that you went to the movies with your friend" [O; failed Di]; Client: [silence]; Therapist: "Sorry, I wonder if you finally got to go with your friend to the movies").

A set of coding criteria, along with examples and nonexamples for each code can be accessed online (http://www.aba-elearning.com/documentos/SISC-INTER.pdf). Interrater agreement was found to be good or excellent in the original study (Froján et al., 2008) with Cohen's κ at 0.68–0.84 and observer accuracy between 87 and 93% (see Bakeman, 2000; Bakeman, Quera, McArthur, & Robinson, 1997, for standard criteria).

Both therapeutic activities and behavior categories were coded continuously (e.g., observation codes were assigned as soon as they could be identified by the observer). A given code entailed a single turn or utterance within a turn from the therapist. In addition, sequential utterances denoting different putative behavior functions could be assigned within a single turn (see examples in the online coding materials). Observers attended a 15-hour workshop by the principal investigator (MX-F) in order to master the observation system (training details and materials are available upon request). For the purposes of analysis, behavior code data were converted into rate per minute data within 10-min intervals.

It should be noted that not all of the SISC-INTER categories are related to a specific behavioral concept (e.g., instructional and informative). It may also be the case that a given verbal utterance could be associated with various behavioral functions (Skinner, 1957). These formal limitations to the validity of the SISC-INTER have been addressed by (a) using nonconstruct-related codes whenever there was no formal evidence of construct-related codes, and (b) assigning the most salient category whenever more than one behavioral function could

apply to a given verbal utterance (see coding rules and examples in the supplementary online material for further details).

Setting and Apparatus

All sessions were videotaped using a camcorder permanently installed in one of the corners of the session room. The camcorder was located behind the participant's back and directed toward the therapist. Observation, coding, and inter- and intrarater agreement analyses were obtained using The Observer® XT (versions 6.0 and 7.0; Noldus Information Technology Inc., Leesburg, Virginia).

Observational Data Analysis

All sessions were divided into 10-min intervals for the purpose of analysis and graphical display. The total number of occurrences of each verbal behavior code was computed for each 10-min interval.

Interobserver Agreement

The senior coinvestigator of this project (MXF-P) provided extensive training on the use of the SISC-INTER to two observers. Interrater agreement was obtained by randomly selecting four full-session recordings that were also coded by a second observer. The second observer underwent the same training in the observation protocol as the primary observer. For the purposes of testing intrarater agreement, once all the sessions were coded, the primary observer recoded one of the sessions. Inter- and intrarater agreement analyses were conducted on a second-by-

second basis but using a tolerance window of ±1 second and calculated in terms of Cohen's κ coefficients.

Known-Group Validity

Known-group validity is a strategy that indirectly assesses the validity of a scale or set of observation codes by demonstrating that the scale's output varies systematically depending upon known performances of the construct that the scale is intended to measure (Netemeyer, Bearden, & Sharma, 2003; Portney & Watkins, 2008). Given that the verbal behavior codes were defined according to putative behavioral functions (e.g., reinforcement), known-group validity analyses were conducted in order to test whether the codes were consistent with the expected performances of the behavioral constructs that they represented. A direct verification of the operant functions involved would have necessitated experimental manipulations that would have been incompatible with an ongoing observation of spontaneous therapist–client interaction. However, known-group validity does provide indirect evidence of the presence of the putative behavioral functions suggested by the number and sequence of the behavior codes.

For the purpose of discriminating between meaningful treatment stages and, thus, the potential differential occurrences of the observation codes, sessions were assigned to two sets of categories: the intervention phase, and the specific therapeutic activity. The intervention phase incorporated sessions primarily focused on assessment, treatment, or follow-up. The therapist determined when the intervention process moved from assessment to treatment and from treatment to follow-up. Specific therapeutic activities were where episodes of salient clinical activity occurred within each session. These categories attempted to capture major forms of clinical interaction in terms of topography and, therefore, with no reference to behavioral processes. Recorded activities were based on a set of four categories developed a priori by senior

behavior therapists (coauthors MXF-P and MM-F): assessing, explaining, Socratic method, and providing clinical guidance. These activities are independent from general intervention stages (i.e., assessment, treatment, follow-up), which were broadly defined by the therapy. Definitions for the activities are presented below:

1. Assessing: The therapist queries the client on events and self-descriptions that may provide useful information with regard to goals and clinical needs and also to establish the effect of the intervention.
2. Explaining: The therapist provides authoritative/technical information on clinically relevant events being discussed with the client.
3. Socratic method: The therapist and client engage in verbal interaction in which the therapist is attempting to provide alternative explanations to the client's descriptions of events on clinically relevant issues. Discussions about dysfunctional thoughts or cognitive restructuring are included in this category.
4. Providing clinical guidance: The therapist discusses information (not necessarily providing instructions) about behavioral repertoires to be acquired by the client in keeping with intervention goals.

Specific activities were coded by two trained external senior therapists who did not have access to the verbal behavior recording protocols and were unaware of the purpose of the study. Observers were simply provided with the operational definitions of these activities. They received training until reaching 90% agreement in coding the therapeutic activities. Codes were assigned continuously and, therefore, any combination of short episodes of any activity could overlap within a given 10-min interval.

The hypotheses below were derived from the applied behavior–analytic literature and, therefore, are consistent with the view that clinical interaction can be conceptualized as an

operant process (Froján-Parga, Calero-Elvira, & Montaño-Fidalgo, 2009; Hamilton, 1988; Skinner, 1988). If the therapist's coded verbal behavior is suggestive of a specific behavioral (operant) function, then the construct validity of the SISC-INTER will be supported.

The descriptive nature of the observation upon which this study is based does not allow one to determine clearly what behavior mechanism, if any, may have been operating. The focus of the study is to design and implement a known-group validity strategy. The possibility exists that the results would not be consistent with the presence of the hypothesized operant processes. Therefore, consistency of the results with the hypothesized processes provides evidence of the presence of those processes. "Consistent with" in this context is of course not equivalent to "caused by" or "mediated by." Therefore, this study only provides an indication that some of the clinical processes analyzed vary consistently with operant processes. Whether those processes were in fact present is for future experimental studies to answer. This approach is of course not new and known-group validity analyses have been used to support the validity of assessment instruments in various clinical disciplines (e.g., Chiang, Hinds, Yeh, & Yang, 2008; Geroldi, Galluzzi, Testa, Zanetti, & Frisoni, 2003).

What follows is a description of the hypotheses we have developed to determine the potential presence of operant processes during clinical interaction. A systematic summary of these hypotheses is presented in Table 2.

Hypothesis 1: Motivational utterances will be higher during periods of clinical guidance. The therapist will use motivational strategies more frequently whenever encouraging specific behavior change strategies.

Hypothesis 2: Reinforcement will occur more frequently during treatment compared to the assessment phase, when active behavior change strategies are being implemented. Behavior therapists make constant use of reinforcement for a number of therapeutic goals

such as shaping adjusted descriptions of feelings, behaviors, and other events, and providing positive feedback on clinically relevant behavior (Kohlenberg & Tsai, 1991). In keeping with the concept of the three-term contingency, it is expected that both reinforcement episodes and also discriminative-reinforcing sequences will be more prevalent during treatment.

Hypothesis 3: The number of informative events and informative–discriminative sequences will be higher during the Socratic dialogue and clinical guidance compared to during other clinical activities. The therapist is expected to provide technical and authoritative information in relation to the client's concerns and when behavior change strategies are being implemented (i.e., Socratic dialogue, clinical guidance). This hypothesis capitalizes on the view of clinical interaction as a sequence of contingencies of reinforcement delivered by the therapist, where informative episodes may be specifying a later client response, in turn signaled by specific discriminative events (Birk, 1968; Callaghan, Naugle, & Follette, 1996; Hamilton, 1988; Rosenfarb, 1992). For example, the therapist may initially point out that exhibiting a specific response within a social context will have a consequence, and this behavior could be said to have an informative function although subsequent presentations of the social context may act as discriminative stimuli for the response. Such informative–discriminative sequences may be characteristic of the "Socratic method" and "providing clinical guidance" activities.

Hypothesis 4: Discriminative and reinforcing events will be more mutually dependent on each other during Socratic dialogue sessions, but not necessarily more frequent. This hypothesis suggests that a shaping process may be taking place during clinical verbal dialogue. Although the therapist may maintain a high rate of reinforcement throughout intervention (e.g., to keep patients motivated, strengthen clinical gains), shaping

requires a close relationship between specific discriminative events and reinforcement. Such a conceptualization of clinical verbal interaction will be consistent with a higher dependence on discriminative and reinforcing events, while their overall level of occurrence may not change. For verbal shaping to occur, natural or artificial contingencies posing consequences for verbal behavior need to be in place. Those contingencies will increase or lower the strength of different verbal classes. There are a number of demonstrations of verbal shaping in the literature (e.g., Greenspoon, 1955; Lovaas, 1964; Rosenfeld & Baer, 1970). Importantly, shaped verbal behavior, once established, may also be included in other verbal processes, such as verbal governance (Catania, 2006), and say–do correspondence (Lima & Abreu-Rodrigues, 2010), which could provide an explanation for outside-session behavior changes mediated by clinical interaction.

Hypothesis 5: Punishing functions will be more frequent during the Socratic method compared to other sections of the intervention process. Consistent with the view of clinical dialogue as a form of verbal shaping (Froján-Parga et al., 2009), it is likely that the therapist will provide a higher amount of feedback during Socratic dialogue, part of which could be negative feedback. Challenging the clients' thoughts and urging them to defend their ideas may be highly aversive or punishing at times (Poppen, 1989). It should be noted, however, that under various scenarios, getting the client to verbalize propositions that in the long term would be better reduced, might be a valuable clinical goal. However, we have framed this hypothesis within episodes of Socratic dialogue because this is the stage in which the therapist is likely to attempt the reduction of targeted propositions directly.

Hypothesis 6: Informative–discriminative sequences will be more prevalent than instructional–discriminative sequences during clinical guidance and Socratic dialogue. Again, this hypothesis is consistent with the view that the client–therapist verbal interchange during clinical guidance and Socratic dialogue can be conceptualized as a process in which the therapist is presenting contingencies of reinforcement. The informative function frequently serves to specify what topography is going to be reinforced, as opposed to a merely instructional process (Birk, 1968).

Hypotheses 1–5 were tested using a time-series ARIMA model. This allows one to identify the impact of an independent variable from single-subject data while controlling for the potential impact of the data trend from the analysis. For the purposes of analyzing Hypothesis 4, the square of the difference of the occurrences of discriminative and reinforcing events (per 10-min session) was computed as a new dependent variable. Lower values of this variable during the Socratic dialogue, as opposed to other stages of the intervention, would provide support for the hypothesis (i.e., that lower values would indicate that unpaired discriminative or reinforcement codes were more unlikely during the Socratic dialogue). Spearman rank correlation coefficients between discriminative and reinforcing events across the Socratic dialogue and all other therapeutic activity sections were also computed with regard to the same hypothesis. Given that in Hypothesis 6, the relative occurrence of two events were compared within specific clinical activities (clinical guidance and Socratic dialogue) rather than temporally, the comparison terms were tested using nonparametric comparisons for correlated samples and not using ARIMA models.

ARIMA models evaluate the influence of exogenous events on the level, variability, and trend of a series of data over time (Aguirre-Jaime, 1994) and involve a two-step process. The first step begins with the identification of suitable ARIMA models; typically models that represent the

preintervention periods prior to the introduction of the independent variable to be tested (Box & Tiao, 1975). The identified model is then reapplied, using the full sample, to test the effects of an independent variable on the level of the time series. Independent variables can be represented as binary variables that indicate either the absence of the variable prior to an event or the presence of the variable during and after an event (McCleary & Hay, 1980). A new variable is added to the model to account for the effect of exogenous intervention (e.g., inception of a specific treatment component). By comparing the level of the postintervention time series to that of the preintervention series, the statistical significance of the effect can be assessed.

Step 1. ARIMA modeling: The general approach of Box-Jenkins ARIMA modeling involves model identification, estimation, and diagnostic checking (Aguirre-Jaime, 1994). During the identification stage, data was visually checked and the Dickey-Fuller test was used to discard stationarity (a dependent variable nondependent on time). If the *t* statistic of the Dickey-Fuller test is positive and nonsignificant ($p > .05$), the process is nonstationary and there is probably a time-dependent trend. A tentative model was then formulated consisting of autoregressive (AR) and moving average (MA) components. The parameters of AR and MA can be adjusted to achieve an optimal weighting of the trend and level components of the data series. AR and MA coefficients ought to be significant and fall within –1 and 1 for the AR and MA coefficients (Aguirre-Jaime, 1995). Finally, to test the goodness of fit of the model to the data, Akaike's Information Criterion and Schwartz's Bayesian Criterion were used. These criteria are interpreted in relative terms: among multiple competing models, the model with smaller values for these criteria is preferred (Aguirre-Jaime, 1994, p. 278).

Step 2. Independent variable: Once a successful ARIMA model is identified, an intervention component was added in the form of a binary variable (i.e., presence/absence of the independent variable of interest). Stationarity of each time series was again checked through the

augmented Dickey-Fuller test. Once a tentative ARIMA model was identified, the intervention models were implemented after inserting each intervention individually. Significant AR and MA coefficients were included to show that data variability, not accounted by the independent variable of interest, was attributable to data noise rather than periodicity or trend (Aguirre-Jaime, 1994, p. 114). Autocorrelation and partial autocorrelation function plots were used to determine the value of AR and MA components. All statistical analyses were conducted with STATA Version 8.1 (Stata Corporation, College Station, Texas).

Data were also summarized graphically using both a cumulative record (for verbal events) and network data graphs (for two-term verbal sequences across intervention phases and clinical activities). NodeXL Version 1.0.1.88 was used for network data graphing (NodeXL Team, 2010). Arrow and node sizes were made proportional to the total number of two-term sequences and the total of behavior code occurrences. Three-term contingency sequences composed of identical codes, and sequences starting with codes representing consequent events (those coded as reinforcing or punishing) were not graphed.

Results

Clinical sessions had an average duration of 50 min each and a total overall duration of 8 hours. Video clips from the sessions comprised forty-seven 10-min periods and generated 4,493 event recordings. Figures 1 and 2 summarize the verbal behavior events and the two-term verbal behavior sequences, respectively. Both inter- and intraobserver agreement estimates were within the range of standard criteria (Cohen's $\kappa = 0.64–0.67$ and 0.76, respectively; see Martínez-Martin, 2010).

Table 2 presents a summary of the hypothesis-driven known-group validity analyses. All dependent variables were stationary (Dickey-Fuller $Z < 0$; MacKinnon $p < .001$). All Z statistics

for AR and MA components were significant or close to statistical significance ($p \leq .1$). Akaike's information criterion and Schwarz's criterion for assessing ARIMA models' goodness of fit were computed whenever several competing models with significant AR and MA components were available. However, AR and partial AR function plots suggested that high-order AR and MA terms were not adequate with the exception of the ARIMA models for Hypothesis 4 (AR[2], MA[2]) and Hypothesis 5 (AR[5], MA[5])—see Table 2.

With regard to Hypothesis 1, whereas motivational events were numerically higher during clinical guidance periods, the ARIMA coefficient showed only a trend toward significance, $\Phi = 2.68$, $p = .068$, 95% CI (–0.20, 5.55). Models computed in support of Hypothesis 2 indicated that reinforcing events were more frequent during the treatment phase, $\Phi = 6.33$, $p < .001$, 95% CI (4.46, 8.20), as were discriminative–reinforcement two-term sequences, $\Phi = 2.45$, $p < .001$, 95% CI (1.50, 3.39). In terms of Hypothesis 3, informative and informative–discriminative sequences occurred more often during Socratic dialogue and clinical guidance, $\Phi = 12.48$, $p < .001$, 95% CI (6.73, 18.23); $\Phi = 3.37$, $p < .001$, 95% CI (1.48, 5.25), respectively. Hypothesis 4 obtained partial support as evidenced by the lack of difference between the occurrence of discriminative–reinforcement sequences across Socratic dialogue and other clinical activities, $\Phi = 0.57$, $p = .698$, 95% CI (–2.31, 3.46). A second analysis, modeling the square of the difference between discriminative and reinforcing events, generated a trend toward significance suggesting that the difference between discriminative and reinforcing events is inferior during Socratic dialogue, $\Phi = -160.96$, $p = .063$, 95% CI (–330.71, 8.79). In addition, the Spearman rank correlation coefficients between discriminative and reinforcing events were also supportive of Hypothesis 4, Socratic method: $r_s = .61$, $p = .006$, 95% CI (.21, .83); other: Socratic method: $r_s = .27$, $p = .159$, 95% CI (–.11, .59). This evidence provides tentative support for the view that discriminative and reinforcing events may be more mutually dependent during Socratic dialogue. With regard to

Hypothesis 5, there was evidence of increased occurrence of punishing functions during periods of Socratic dialogue, $\Phi = 1.99$, $p = .007$, 95% CI (0.55, 3.43). Finally, nonparametric comparisons for correlated samples provided support for Hypothesis 6. Specifically, the difference between informative–discriminative and instructional–discriminative sequences was deemed significant during both Socratic dialogue, Wilcoxon $Z = -3.07$, $p = .002$, and clinical guidance activities, Wilcoxon $Z = -4.03$, $p < .001$ (see Table 2).

Discussion

The SISC-INTER was designed to be used as an observation protocol for the descriptive analysis of a therapist's verbal behavior. It does not, per se, provide solid evidence for the behavioral functions of the verbal behavior codes being used; this can only be established experimentally. However, a hypothesis-based approach to construct validity (i.e., known-group validity) may constitute a methodological strategy to support the validity of the SISC-INTER. Specifically, the correspondence between the expected occurrences of putative verbal operants and the verbal behavior codes used in the observation protocol were examined. We attempted to provide support for the construct validity of the SISC-INTER by testing a set of hypotheses based on known properties of verbal operants and other elements of verbal behavior that are defined in the SISC-INTER. In addition, the interobserver agreement of the SISC-INTER was examined.

Our results are consistent with a high rate of reinforcing events and discriminative–reinforcement sequences during the treatment phase of clinical interactions. Known-group validity results were also consistent with the presence of operant processes during clinical interaction (Froján-Parga et al., 2009). More specifically, our results suggest that informative–discriminative reinforcement sequences are frequent during the active intervention phases (Socratic method, clinical guidance). Such a sequence of events may be characteristic of verbal

shaping (Hamilton, 1988). In addition, the results also suggested that putative discriminative and reinforcing events were more mutually dependent during periods of Socratic method, commensurate with the concept of (increased) stimulus control of client responses during these periods. The tests for Hypotheses 5 and 6 were also consistent with the conceptualization of aspects of clinical interaction being consistent with operant processes. For example, punishing events were higher during periods of Socratic dialogue compared to other sections of the intervention process. Finally, neither Socratic dialogue nor clinical guidance activities were characterized by direct instruction delivery. Instead, informative–discriminative sequences (noninstructional information exchanges) were frequently presented before discriminative reinforcement sequences. Additional studies are needed to determine the extent to which our observation codes faithfully portray actual behavior processes and if they operate as active components that account for the effects of psychotherapy.

The level of interobserver agreement found was acceptable but not optimal. A potential way to improve it would require the development of observation training materials. Providing specific and immediate feedback to prospective observers has been found highly effective to improve the observation accuracy of health professionals (Goldberg, 2009). Training could be delivered through a series of computer-administered observation problems covering most scenarios under which a given observation code would be assigned.

There are several limitations to our study. First, analyses were restricted to a single case. Future studies may incorporate multiple subjects in order to provide a wider description of the psychometric properties of the SISC-INTER. Second, given that the participant's behavior was not recorded, the client–therapist interaction events could not be analyzed directly and the client's verbal behavior was inferred from analysis of discriminative reinforcement sequences (e.g., the

client's behavior was assumed to be interspersed between discriminative reinforcement sequences).

It should be noted that clinical activities did not always follow intervention phases (assessment, treatment, follow-up). Activities more typical of the treatment sessions were also seen during the assessment phase and vice versa (see Figure 1). Moreover, "assessing," as a specific clinical activity, was present throughout the whole process although the assessment goal varied (e.g., the reason for the referral, effects and issues resulting from treatment implementation, the client's perception of therapeutic achievements). Although the assessment-treatment distinction may provide an informative context in which to evaluate the verbal categories, analyzing the verbal code distributions in terms of the specific momentary goals of the therapist is likely to allow a more sensitive and molecular analysis.

In order to maximize the construct validity of the observation codes suggestive of behavioral function (discriminative, reinforcing, punishing, and motivational), we have added other codes with no parallel behavioral function (informative, instructional). This approach to observing verbal behavior may have strengthened the validity of the codes with a putative behavioral function by limiting the number of verbal events without a clear function that would have been assigned to functional categories.

Our results are descriptive in nature. Analysis of the behavioral categories do not demonstrate that the verbal behavior of the therapist has a specific functional effect such as reinforcement of specific behaviors (cf. Eide, Quera, Graugaard, & Finset, 2004; Follette, Naugle, & Callaghan, 1996; Truax, 1966). The nature of the analysis herein presented is sufficiently molar to allow various interpretations in terms of the potential mechanisms involved, behavioral or otherwise. Within a behavioral framework, our results are consistent with the existence of operant processes during clinical interaction, which fulfills the purposes of a known-

group validity analysis. It remains for future experimental studies to pinpoint specific operant processes. We have mentioned verbal shaping as a potential mechanism involved in clinical interaction (cf. Hypothesis 4). Verbal shaping has been used not only to encourage individuals to utter more frequently specific types of nouns, pronouns, verbs, or semantic categories (Greenspoon, 1955; Mandler & Kaplan, 1956; Taffel, 1953; Zoltan-Ford, 1991) but also to show self-acceptance (Nuthmann, 1957), or to conform to the attitude of other people (Hildum & Brown, 1956). In addition, there are conceptual elaborations suggesting how shaped verbal behavior may lead to behavior change outside the clinical session (Catania, 2003, 2006).

Alternative cognitive hypotheses are compatible with our results. However, Socratic debate has not been described in terms of the sequences of operationally described verbal episodes used in this study. Therefore, the value of alternative hypotheses based on paradigms that are incommensurable with the very definition of the observation codes is uncertain (cf. Oberheim & Hoyningen-Huene, 2010). A given sequence of verbal events could be described both as restructuring thoughts or verbal shaping, but what matters for our present purpose is to what extent our data is consistent with the operant approach. The present analysis was designed to determine to what extent sequences of verbal events vary consistently with known attributes of operant processes.

The hypothesis-driven known-group validity analysis suggests that the set of verbal codes developed by our team can be used to reliably record aspects of the therapist's verbal behavior that are suggestive of specific putative behavioral functions. On another note, various studies have shown the lack of correspondence between descriptive and experimental functional analyses of behavior (Hall, 2005; Thompson & Iwata, 2007). However, the convergence between the descriptive and functional analyses of verbal behavior in typically developed adults has not yet been explored. This study suggests that a novel integration of methods—the combination of the

known-group validity analyses with a behavioral coding system—facilitates the application of a coding system to the therapeutic encounter.

Although this study is a methodological analysis of the SISC-INTER, it suggests exciting future applications. If effective components of therapeutic interaction could be identified reliably, they could be used as quantitative indicators of an ongoing therapeutic intervention. This information would help practitioners to adjust their performance based on reliable data that is able to reflect the extent to which the therapist is using effective treatment components (e.g., reinforcement). Alternative uses would be treatment fidelity monitoring, and applications for professional training and feedback.

Future studies may deal effectively with the shortcomings described above by incorporating any of the following strategies: (a) conducting observations of multiple therapist–client dyads; (b) analyzing the verbal behavior of both the therapist and the client; and (c) developing experimental approaches that test the extent to which specifically defined codes, which are intended to parallel a behavioral function, vary systematically, in conjunction with associated antecedents and consequences. The latter represents a criterion validity strategy that would add significantly to the evidence presented in the current study.

Conclusion

Although the generalizability of the results is limited, this study illustrates a methodological approach to test the reliability and validity of an observation protocol for the analysis of a therapist's in-session verbal behavior from an operant perspective. The application of the SISC-INTER was able to generate high levels of interobserver agreement. More importantly, the results provided support for a set of hypotheses that tested the extent to which verbal behavior codes were indicative of their respective putative behavioral functions.

While process research is a minor field within clinical psychology, its relevance in advancing theory and practice is being increasingly acknowledged (Kazdin, 2008; Thorn, 2007). However, there is a paucity of effective methods to analyze in-session clinical processes. Methodological approaches such as the one suggested in the present report may be useful for the advancement of our understanding of clinical phenomena.

References

Aguirre-Jaime, A. (1995). *Introducción al tratamiento de series temporales: Aplicación a las ciencias de la salud* [Introduction to time-series analysis: Application to health sciences]. Madrid, Spain: Diaz de Santos.

Bakeman, R. (2000). Behavioural observation and coding. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 138–159). Cambridge, UK: Cambridge University Press.

Bakeman, R., Quera, V., McArthur, D., & Robinson, B. F. (1997). Detecting sequential patterns and determining their reliability with fallible observers. *Psychological Methods, 2*, 357-370. doi:10.1037/1082-989X.2.4.357

Birk, L. (1968). Social reinforcement in psychotherapy. *Integrative Psychological and Behavioral Science, 3*, 116-123. doi:10.1007/BF03001142

Borckardt, J. J., Nash, M. R., Murphy, M. D., Moore, M., Shaw, D., & O'Neil, P. (2008). Clinical practice as natural laboratory for psychotherapy research: A guide to case-based time-series analysis. *American Psychologist, 63*, 77-95. doi:10.1037/0003-066X.63.2.77

Box, G. E. P., & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of American Statistical Association, 70*, 70–79.

Busch, A. M., Kanter, J. W., Callaghan, G. M., Baruch, D. E., Weeks, C. E., & Berlin, K. S. (2009). A micro-process analysis of functional analytic psychotherapy's mechanism of change. *Behavior Therapy, 40*, 280-290. doi:0005-7894/08/0280–0290$1.00/0

Callaghan, G. M., Naugle, A. E., & Follette, W. C. (1996). Useful constructions of the client–therapist relationship. *Psychotherapy: Theory, Research, Practice, Training, 33*, 381-390. doi:10.1037/0033-3204.33.3.381

Catania, A. C. (2003). Verbal governance, verbal shaping, and attention to verbal stimuli. In K. A. Lattal & P. H. Chase (Eds.), *Behavior theory and philosophy* (pp. 301–321). New York: Kluwer.

Catania, A. C. (2006). Antecedents and consequences for words. *Analysis of Verbal Behavior, 22,* 89-100.

Chiang, Y. C., Hinds, P. S., Yeh, C. H., & Yang, C. P. (2008). Development and psychometric testing of a Chinese version of the Fatigue Scale–Children in Taiwan. *Journal of Clinical Nursing, 17,* 1201-1210. doi:10.1111/j.1365-2702.2007.02138.x

Dougher, M. J. (1989). A functional analysis of a behavior analyst's functional analysis. *Analysis of Verbal Behavior, 7,* 19-23.

Eide, H., Quera, V., Graugaard, P., & Finset, A. (2004). Physician–patient dialogue surrounding patients' expression of concern: Applying sequence analysis to RIAS. *Social Science and Medicine, 59,* 145-155. doi:10.1016/j.socscimed.2003.10.011

Epstein, N. B., & Baucom, D. H. (2002). *Enhanced cognitive–behavioral therapy for couples: A contextual approach.* Washington, DC: American Psychological Association.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: Wiley.

Follette, W. C., Naugle, A. E., & Callaghan, G. M. (1996). A radical behavioural understanding of the therapeutic relationship in effecting change. *Behavior Therapy, 27,* 623-641. doi:10.1016/S0005-7894%2896%2980047-5

Froján, M. X., Montaño, M., Calero, A., García, A., Garzón, A., & Ruiz, E. (2008). Sistema de categorización de la conducta verbal del terapeuta [A coding system for the therapist's verbal behavior]. *Psicothema, 20,* 603-609.

Froján-Parga, M. X., Calero-Elvira, A., & Montaño-Fidalgo, M. (2009). Analysis of the therapist's verbal behavior during cognitive restructuring debates: A case study. *Psychotherapy Research, 24,* 1-12. doi:10.1080/10503300802326046

Geroldi, C., Galluzzi, S., Testa, C., Zanetti, O., & Frisoni, G. B. (2003). Validation study of a CT-based weighted rating scale for subcortical ischemic vascular disease in patients with mild cognitive deterioration. *European Neurology, 49,* 193-209. doi:10.1159/000070183

Goldberg, D. (2009). Training family doctors to recognize psychiatric illness with increased accuracy. *Lancet, 316,* 521-523.

Greenspoon, J. (1955). The reinforcing effect of two spoken sounds on the frequency of two responses. *American Journal of Psychology, 68,* 409-416. doi:10.2307/1418524

Hall, S. S. (2005). Comparing descriptive, experimental and informant-based assessments of problem behaviors. *Research in Developmental Disabilities, 26,* 514–526. doi:10.1016/j.ridd.2004.11.004

Hamilton, S. A. (1988). Behavioral formulations of verbal behavior in psychotherapy. *Clinical Psychology Review, 8,* 181-193. doi:10.1016/0272-7358(88)90058-X

Hayes, S. C., Blackledge, J. T., & Barnes-Holmes, D. (2001). Language and cognition: Constructing an alternative approach within the behavioral tradition. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational frame theory: A post-Skinnerian account of human language and cognition* (pp. 3–20). New York: Plenum Press.

Hildum, D., & Brown, R. W. (1956). Verbal reinforcement and interviewer bias. *Journal of Abnormal and Social Psychology, 53,* 108-111. doi:10.1037/h0040518

Jones, E. E., Ghannam, J., Nigg, J. T., & Dyer, J. F. (1993). A paradigm for single-case research: The time series study of a long-term psychotherapy of depression. *Journal of Consulting and Clinical Psychology, 61,* 381-394. doi:10.1037/0022-006X.61.3.381

Kazdin, A. E. (2008). Evidence-based treatment and practice. New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist, 63,* 146-159. doi:10.1037/0003-066X.63.3.146

Kohlenberg, R. J., & Tsai, M. (1991). *Functional analytic psychotherapy: Creating intense and curative therapeutic relationships.* New York: Plenum.

Lambert, M. J., & Hill, C. E. (1994). Assessing psychotherapy outcomes and processes. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 72–113). New York: Wiley.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics 33*, 159–174. doi:10.2307/2529310

Lima, E. L., & Abreu-Rodrigues, J. (2010). Verbal mediating responses: Effects on generalization of say–do correspondence and noncorrespondence. *Journal of Applied Behavior Analysis, 43,* 411-424.

Lovaas, O. I. (1964). Cue properties of words: The control of operant responding by rate and content of verbal operants. *Child Development, 35,* 245–256. doi:10.2307/1126587

Madler, G., & Kaplan, W. K. (1956). Subjective evaluation and reinforcing effect of a verbal stimulus. *Science, 124,* 582-583.

Martínez-Martín, P. (2010). Composite rating scales. *Journal of the Neurological Sciences, 289,* 7-11. doi:10.1016/j.jns.2009.08.013

McCleary, R., & Hay, R. (1980). *Applied time series analysis*. Beverly Hills, CA: Sage.

Moore, J. (1991). A retrospective appreciation of Willard Day's contributions to radical behaviorism and the analysis of verbal behavior. *Analysis of Verbal Behavior, 9,* 97-104.

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and applications*. Thousand Oaks, CA: Sage.

Node XL Team, Microsoft Corporation. (2010). *NodeXL: Network overview, discovery and exploration for Excel*, version 1.0.1.113. Retrieved from http://www.codeplex.com/nodexl.

Nuthmann, A. M. (1957). Conditioning of a response class on a personality test. *Journal of Abnormal and Social Psychology, 54*, 19-23. doi:10.1037/h0044808

Oberheim, E., & Hoyningen-Huene, P. (2010). The incommensurability of scientific theories. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2010 ed.). Retrieved from http://plato.stanford.edu/archives/fall2010/entries/incommensurability.

Poppen, R. L. (1989). Some clinical implications of rule-governed behavior. In S. C. Hayes (Ed.), *Rule-governed behavior: Cognition, contingencies and instructional control* (pp. 325–357). New York: Plenum Press.

Portney, L. G., & Watkins M. (2008). *Foundations of clinical research: Applications to practice* (3rd ed). Upper Saddle River, NJ: Prentice-Hall.

Rosenfarb, I. S. (1992). A behaviour analytic interpretation of the therapeutic relationship. *Psychological Record, 42*, 341-354.

Rosenfeld, H. M., & Baer, D. M. (1970). Unbiased and unnoticed verbal conditioning: The double agent robot procedure. *Journal of the Experimental Analysis of Behavior, 14*, 99–107.

Simon, A., & Agazarian, Y. (2000). The System for Analyzing Verbal Interaction. In A. P. Beck & C. M. Lewis. (Eds.), *The process of group psychotherapy: Systems for analyzing change* (pp. 357–380). Washington, DC: American Psychological Association.

Skinner, B. F. (1957). Multiple causation. In *Verbal Behavior* (chap. 9, pp. 227–252). New York: Appleton-Century-Crofts.

Skinner, B. F. (1988). The operant side of behavior therapy. *Journal of Behavior Therapy and Experimental Psychiatry, 19*, 171-179. doi:10.1016/0005-7916%2888%2990038-9

Smith, J. D., Wolf, N. J., Handler, L., & Nash, M. R. (2009). Testing the effectiveness of family therapeutic assessment: A case study using a time-series design. *Journal of Personality Psychology, 91*, 518-536. doi:10.1080/00223890903228331

Taffel, C. (1955). Anxiety and the conditioning of verbal behavior. *Journal of Abnormal and Social Psychology, 51*, 496-501. doi:10.1037/h0040463

Thompson, R. H., & Iwata, B. A. (2007). A comparison of outcomes from descriptive and functional analyses of problem behavior. *Journal of Applied Behavior Analysis, 40*, 333-338. doi:10.1901/jaba.2007.56-06

Thorn, B. E. (2007). Evidence-based practice in psychology. *Journal of Clinical Psychology, 63*, 607-609. doi:10.1002/jclp.20384

Truax, C. B. (1966). Reinforcement and nonreinforcement in Rogerian psychotherapy. *Journal of Abnormal Psychology, 71*, 1-9. doi:10.1037/h0022912

Zoltan-Ford, E. (1991). How to get people to say and type what computers can understand. *International Journal of Man-Machine Studies, 34*, 527-547. doi:10.1016/0020-7373(91)90034-5

Table 1 Qualitative Summary of Sessions' Topics

Session	Phase	Period	Session Topics
1	AS	1–6	*Major:* Work plan and shaping of realistic expectancies about treatment. *Minor:* Homework and revisions.
2	AS	7–11	*Major:* Reason for referral. Hypotheses about client's couple conflict and discussion of strategies for change. *Minor:* Homework and revisions, chat.
3	AS	12–17	*Major:* Discussion of client's dysfunctional verbalizations associated with a recent distressing event (Socratic dialogue). *Minor:* Discussion on how to behave in the event of couple conflict episodes.
4	TT	18–21	*Major:* Interview of client's partner. Presentation of case conceptualization (analogue functional analysis). *Minor:* Clarification on intervention organization, homework and revisions.
5	TT	22–26	*Major:* Presentation of behavior change strategies. Discussion of adaptive daily life repertoires. *Minor:* Homework and revisions.
6	TT	27–30	*Major:* Discussion on the benefits of cognitive change. Discussion of adaptive daily life repertoires. *Minor:* Presentation of clinical goals and strategies. Homework and revisions.
7	TT	31–35	*Major:* Discussion of client's verbalizations (Socratic dialogue). Discussion of adaptive daily life repertoires. *Minor:* Discussion of couple difficulties. Clarifications on intervention organization.
8	TT	36–38	*Major:* Assessment of behavior changes achieved so far. Presentation of behavior change strategies. *Minor:* Homework and revisions, chat, clarification on intervention organization.
9	TT	39–44	*Major:* Discussion on the recent changes in client's environment. Discussion on the practical difficulties posed by the therapeutic plan. *Minor:* Assessment of behavior changes achieved so far. Discussion on specific maladaptive thoughts. Presentation of behavior change strategies, chat.
10	FU	45–47	*Major:* Assessment of clinical achievements and revision of intervention strategies for behavior change. *Minor:* Chat.

Note. AS = assessment; Chat = clinically irrelevant conversation; FU = follow-up; Homework and revisions = discussion of intersession behavior and assignments; Period = 10-min period; TT = treatment.

Table 2 Summary of Known-Group Validity Test Results

Hypotheses[a]	Comparison Terms	*N*	*M* ± *SD*	ARIMA Models: Φ ± *SE*[b]	ARIMA Models: AR, MA	Test: Statistic	Test: *p*	Validity Supported
1								
H_0: $Mt_{CG} = Mt_{other}$	Mt_{CG}	23	6.61 ± 5.38	2.68 ± 1.47	1, 1	1.82	.068	Partially
H_1: $Mt_{CG} > Mt_{other}$	Mt_{other}	24	3.58 ± 4.91	(–0.20, 5.55)				supported
2								
H_0: $Rf_{AS} = Rf_{TT}$	Rf_{EV}	21	17.95 ± 7.09	6.33 ± 0.96	1, 1	6.62	< .001	Supported
H_1: $Rf_{AS} < Rf_{TT}$	Rf_{TT}	26	23.38 ± 7.52	(4.46, 8.20)				
H_0: $Di–Rf_{AS} = Di–Rf_{TT}$	$Di–Rf_{EV}$	21	8.81 ± 4.70	2.45 ± 0.48	1, 1	5.07	< .001	
H_1: $Di–Rf_{AS} < Di–Rf_{TT}$	$Di–Rf_{TT}$	26	11.08 ± 4.22	(1.50, 3.39)				
3								
H_0: $If_{SM, CG} = If_{other}$	$If_{SM, CG}$	35	28.83 ± 11.90	12.48 ± 2.91	1, 1	4.26	< .001	Supported
H_1: $If_{SM, CG} > If_{other}$	If_{other}	12	16.50 ± 8.06	(6.73, 18.23)				
H_0: $If–Di_{SM, CG} = If–Di_{other}$	$If–Di_{SM,CG}$	35	7.17 ± 3.14	3.37 ± 0.96	1, 1	3.50	< .001	
H_1: $If–Di_{SM, CG} > If–Di_{other}$	$If–Di_{other}$	12	4.08 ± 3.15	(1.48, 5.25)				
4								
H_0: $Di–Rf_{SM} = Di–Rf_{other}$	$Di–Rf_{SM}$	19	10.42 ± 5.07	0.57 ± 1.47	2, 2	0.39	.698	Partially
H_1: $Di–Rf_{SM} \neq Di–Rf_{other}$	$Di–Rf_{other}$	28	9.82 ± 4.22	(–2.31, 3.46)				supported
H_0: $(Di_{SM}–Rf_{SM})^2 = (Di_{other}–Rf_{other})^2$	$(Di_{SM}–Rf_{SM})^2$	19	109.95 ± 193.80	–160.96 ± 86.61	2, 2	–1.86	.063	
H_1: $(Di_{SM}–Rf_{SM})^2 < (Di_{other}–Rf_{other})^2$	$(Di_{other}–Rf_{other})^2$	28	234.89 ± 325.12	(–330.71, 8.79)				
5								
H_0: $Pn_{SM} = Pn_{other}$	Pn_{SM}	19	3.11 ± 3.00	1.99 ± 0.74	5, 5	2.70	.007	Supported
H_1: $Pn_{SM} > Pn_{other}$	Pn_{other}	28	1.14 ± 1.53	(0.55, 3.43)				

Table 2 *(cont'd)*

Hypotheses[a]	Comparison Terms	*N*	*M* ± *SD*	ARIMA Models Φ ± *SE*[b]	ARIMA Models AR, MA	Test Statistic	Test *p*	Validity Supported
6								
H_0: If-Di_{SM} = It-Di_{SM}	If–Di_{SM}	19	6.63 ± 2.73	-	-	–3.07	.002	Supported
H_1: If-Di_{SM} > It-Di_{SM}	It–Di_{SM}	19	1.16 ± 1.50					
H_0: If-Di_{CG} = It-Di_{CG}	If–Di_{CG}	23	7.70 ± 3.47	-	-	–4.03	< .001	
H_1: If-Di_{CG} > It-Di_{CG}	It–Di_{CG}	23	1.65 ± 1.72					

Note. SE; AR = autoregression coefficients; MA = moving average coefficients; Rf = reinforcement function; Di = discriminative function; Mt = motivational function; If = informative function; It = instructional function; Pn = punishment function; SM = Socratic method; CG = clinical guidance; AS = assessment phase; TT = treatment phase. All dependent variables were stationary (Dickey-Fuller $Z < 0$; MacKinnon $p < .001$). All *Z* statistics for AR and MA components were significant or close to statistical significance ($p \leq .1$). Standardized Z for ARIMA regression and Wilcoxon Z for correlated samples were used as contrast statistics for Hypotheses 1–5 and 6, respectively.

[a]Subscripts refer to independent variable conditions (general and specific clinical activities).

[b]95% confidence intervals in square brackets.

FIGURE CAPTIONS

FIGURE 1 Cumulative record of verbal behavior codes across general and specific clinical activities. Gray areas highlight 10-min periods with the presence of each specific clinical activity (assessment, explanation, Socratic dialogue, and clinical guidance). Codes were assigned continuously, therefore, any combination of behavior codes and activities within a given 10-min interval was possible.

FIGURE 2 Network graph for two-term verbal code sequences across intervention phases and clinical activities. Sequences composed of identical codes or beginning with a consequent event were omitted. Arrow widths are proportional to the total number of two-term sequences. Node size is proportional to the total number of behavior code occurrences. Rf = reinforcing; Di = discriminative; Mt = motivational; If = informative; It = instructional; Pn = punishing.

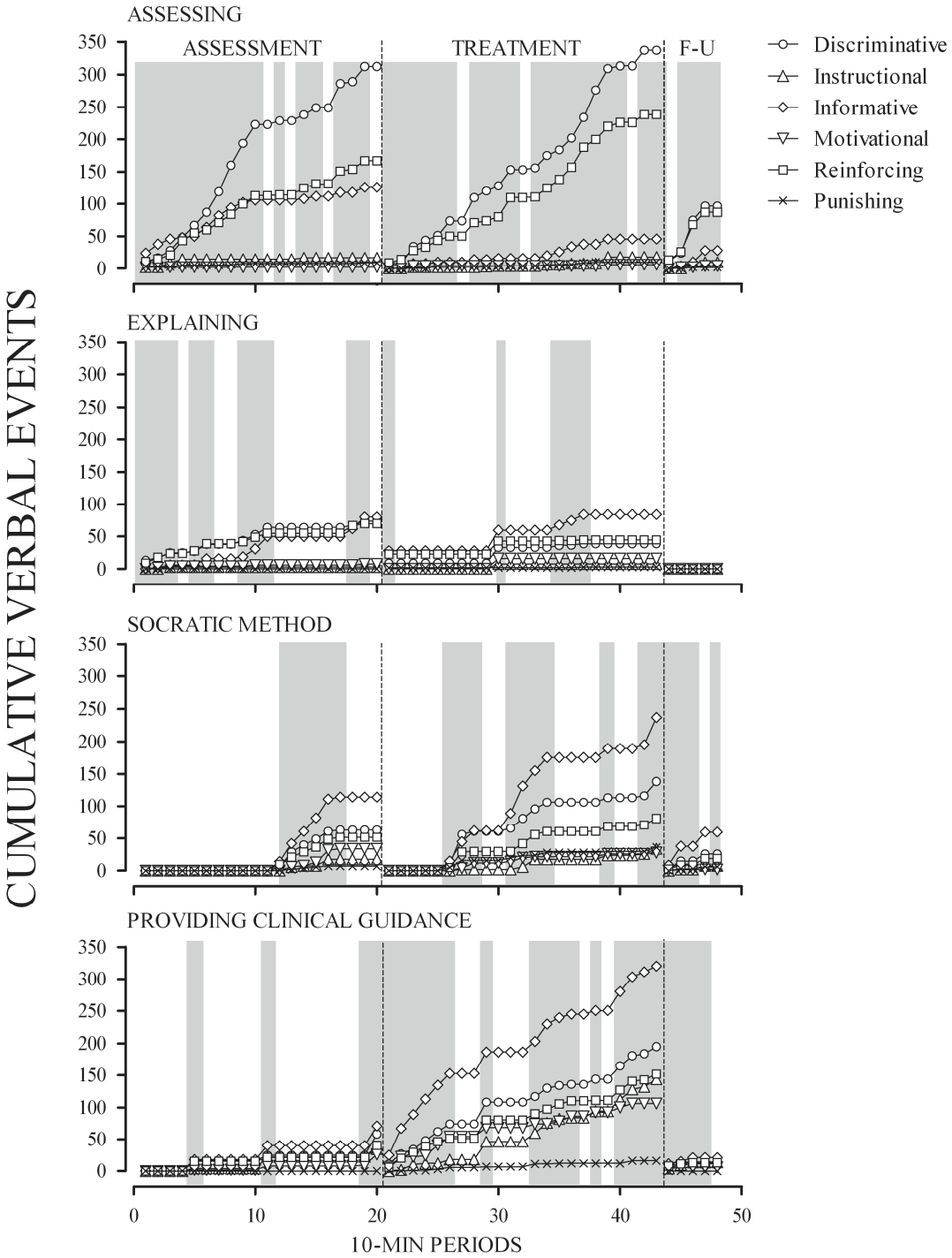

Figure 1

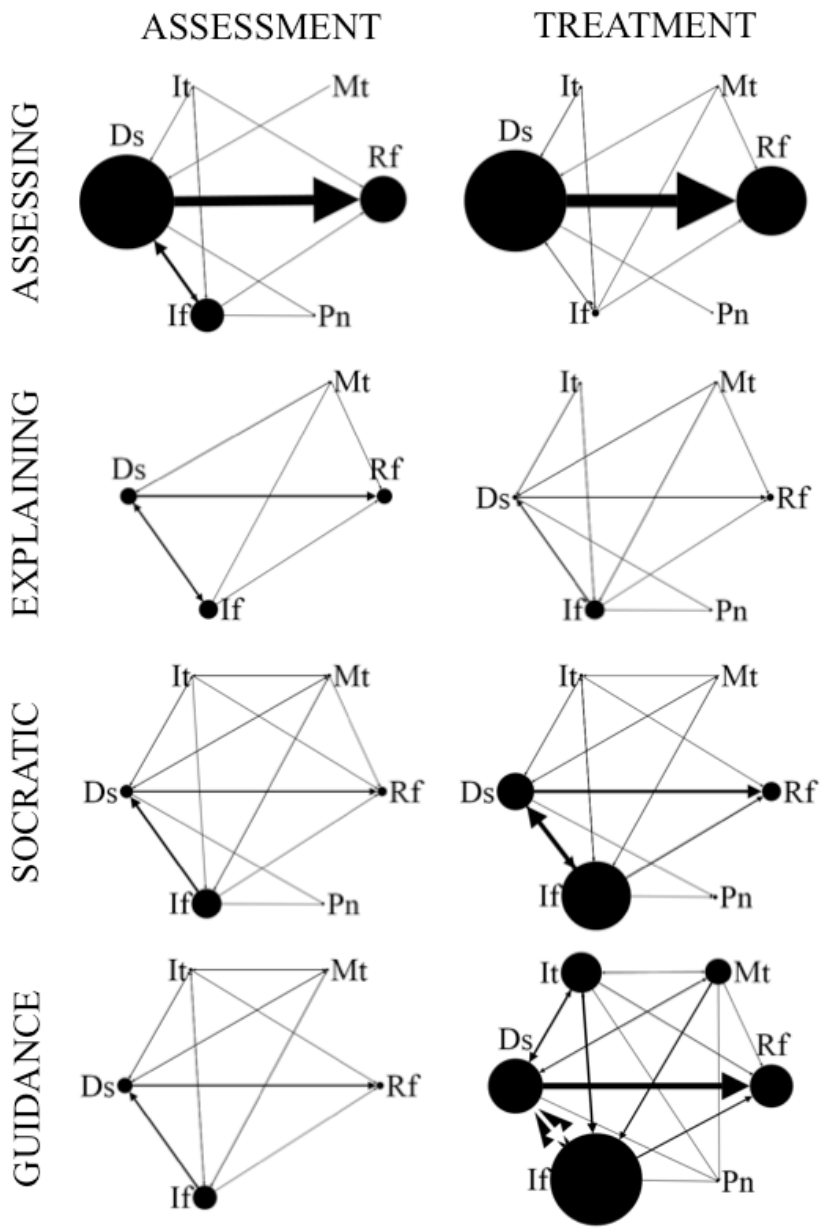

Figure 2